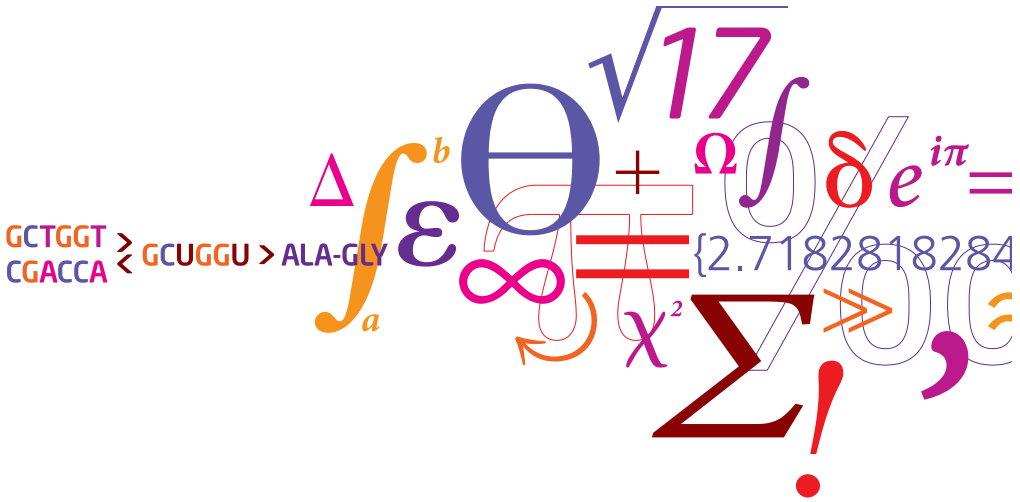
Variant Effect Prediction

Jose MG Izarzugaza, PhD
Integrative Systems Biology Group
Center Biological Sequence Analysis (CBS)



DTU Systems Biology
Department of Systems Biology

Understanding variants in *-omics* times

Traditionally

1 Mutation

=

1 Disease

Phenotype

Function

Mechanism

Lots of hard work

Now (High Throughput Sequencing, NGS)

X Mutations

In

Y Patients

And

Z Conditions

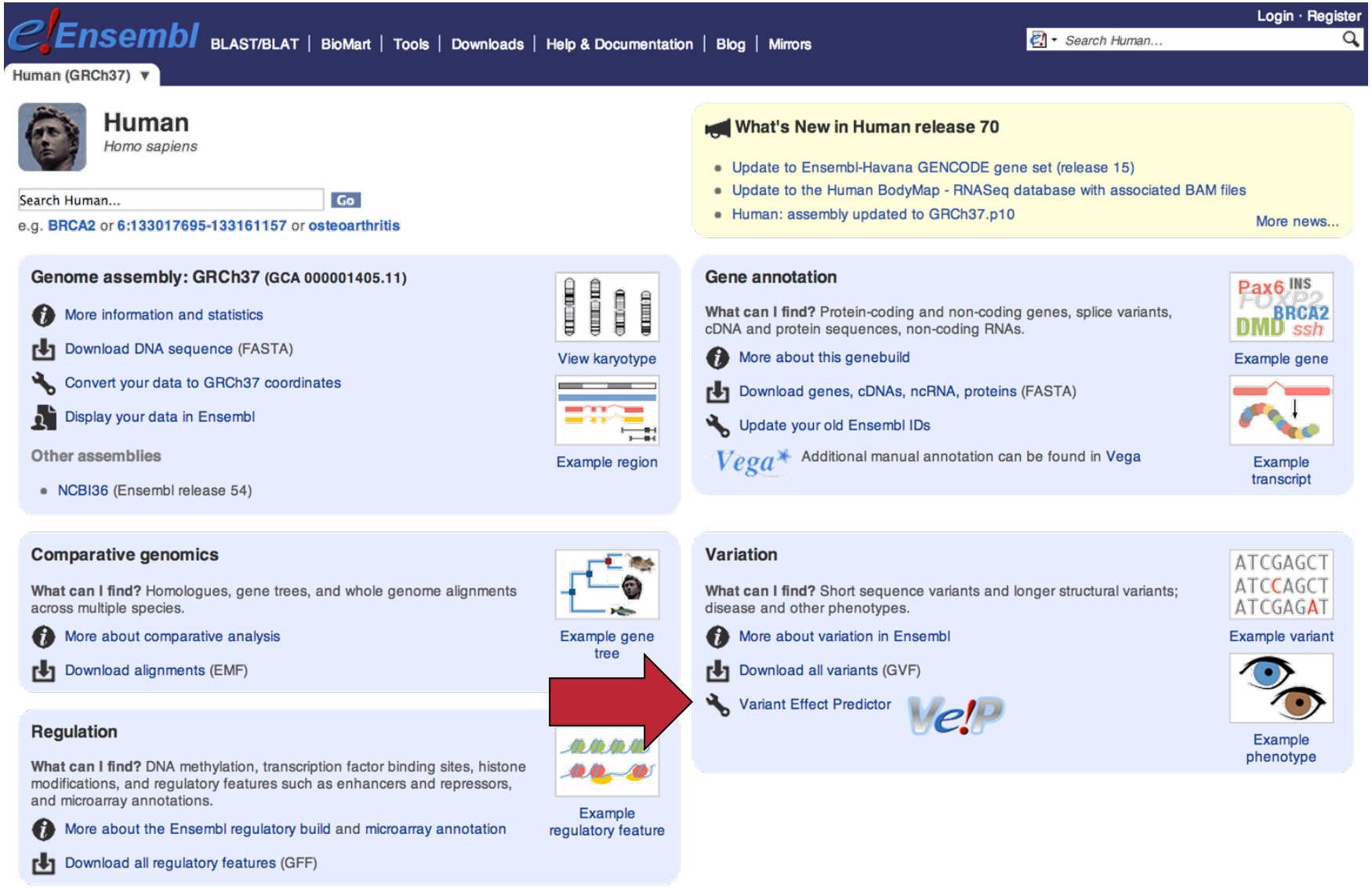
Prediction of

Pathogenicity / Unfeasible

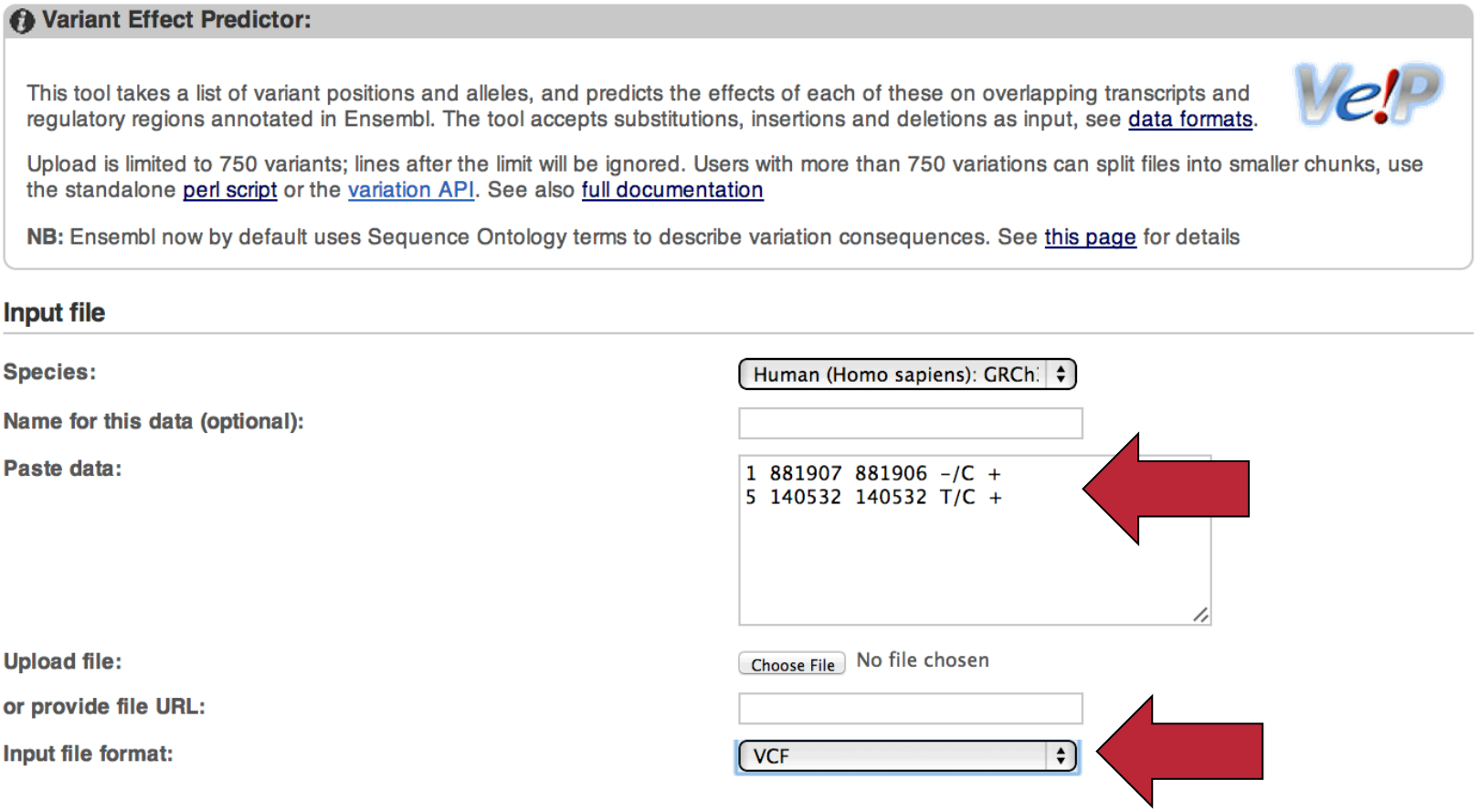
Prioritization

Ensembl Variant Effect Predictor

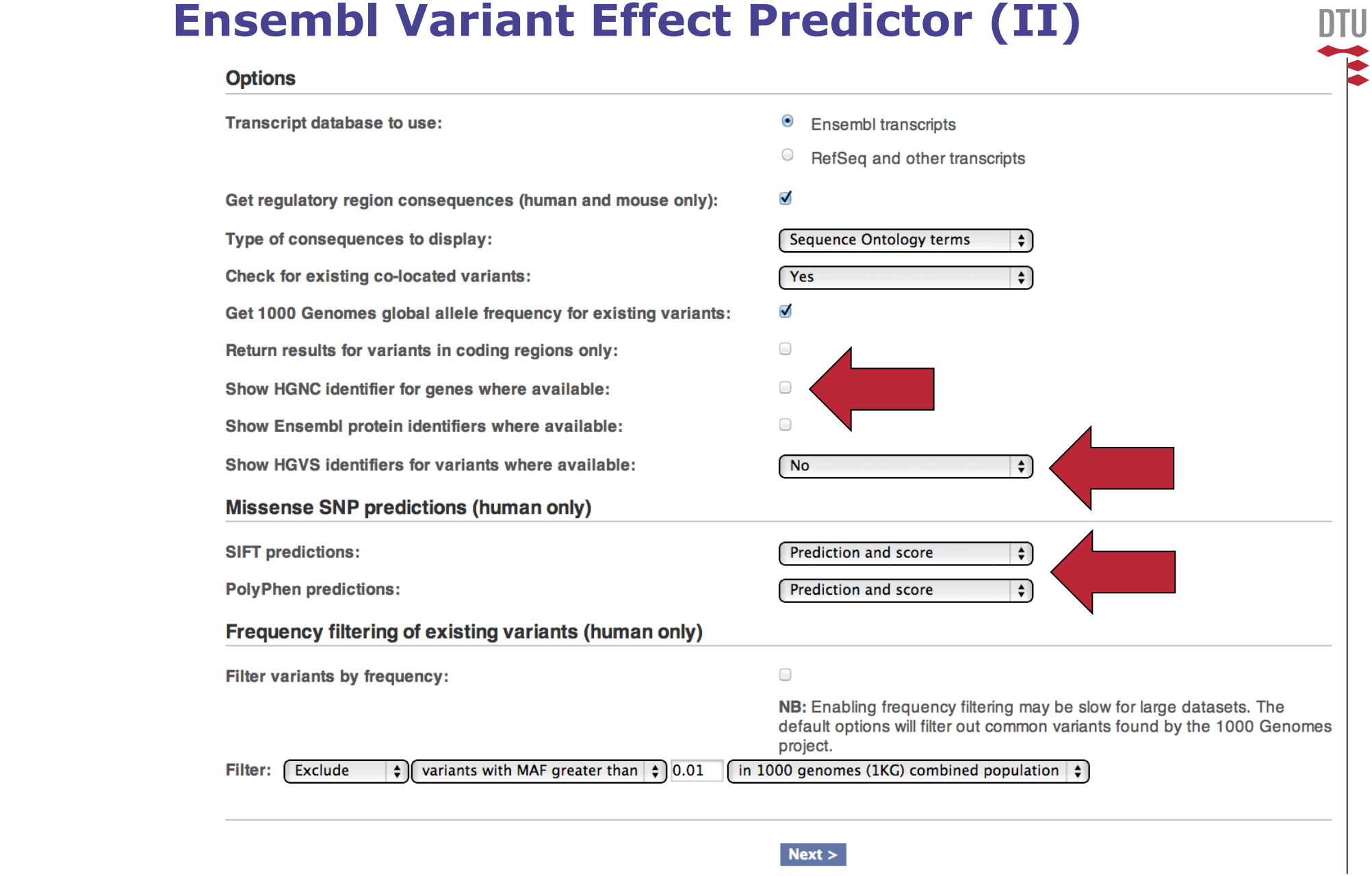
http://www.ensembl.org/Homo_sapiens/Info/Index



Ensembl Variant Effect Predictor (I)



Ensembl Variant Effect Predictor (II)



Ensembl Variant Effect Predictor (Results)

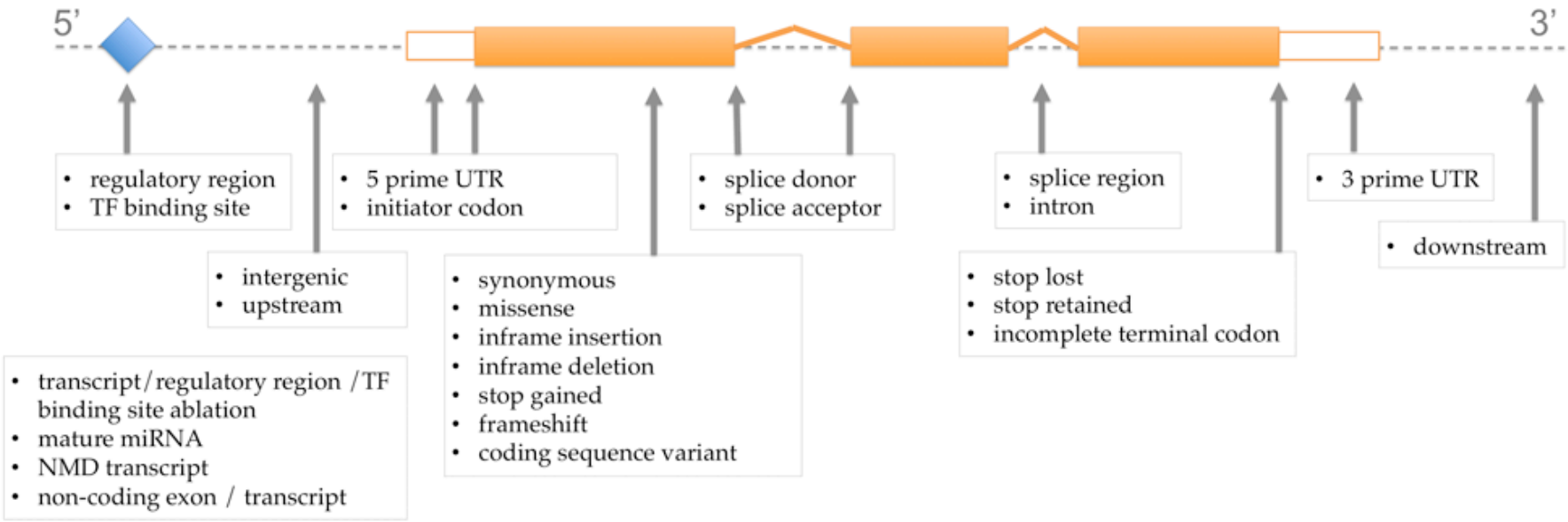
Variant Effect Predictor Results:

Download text version

Show 10 entries | Show/hide columns | Filter

Uploaded Variation	Location	Allele	Gene	Feature	Feature type	Consequence	Position in cDNA	Position in CDS	Position in protein	Amino acid change	Codon change	Co-located Variation	Extra
1_881907_-/C	1:881906-881907	C	ENSG00000187634	ENST00000466827	Transcript	downstream_gene_variant	-	-	-	-	-	-	**DISTANCE**=3724
5_140532_T/C	5:140532	C	ENSG00000249430	ENST00000512035	Transcript	downstream_gene_variant	-	-	-	-	-	rs12516846	**DISTANCE**=554; **GMAF**=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000199540	ENST00000362670	Transcript	downstream_gene_variant	-	-	-	-	-	rs12516846	**DISTANCE**=3670; **GMAF**=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000153404	ENST00000283426	Transcript	missense_variant	160	110	37	V/A	gTa/gCa	rs12516846	**PolyPhen**=benign(0); **SIFT**=tolerated(1); **GMAF**=C:0.1534
5_140532_T/C	5:140532	C	ENSG00000153404	ENST00000502646	Transcript	upstream_gene_variant	-	-	-	-	-	rs12516846	**DISTANCE**=149; **GMAF**=C:0.1534

Showing 11 to 15 of 15 entries | << < 1 2 > >>

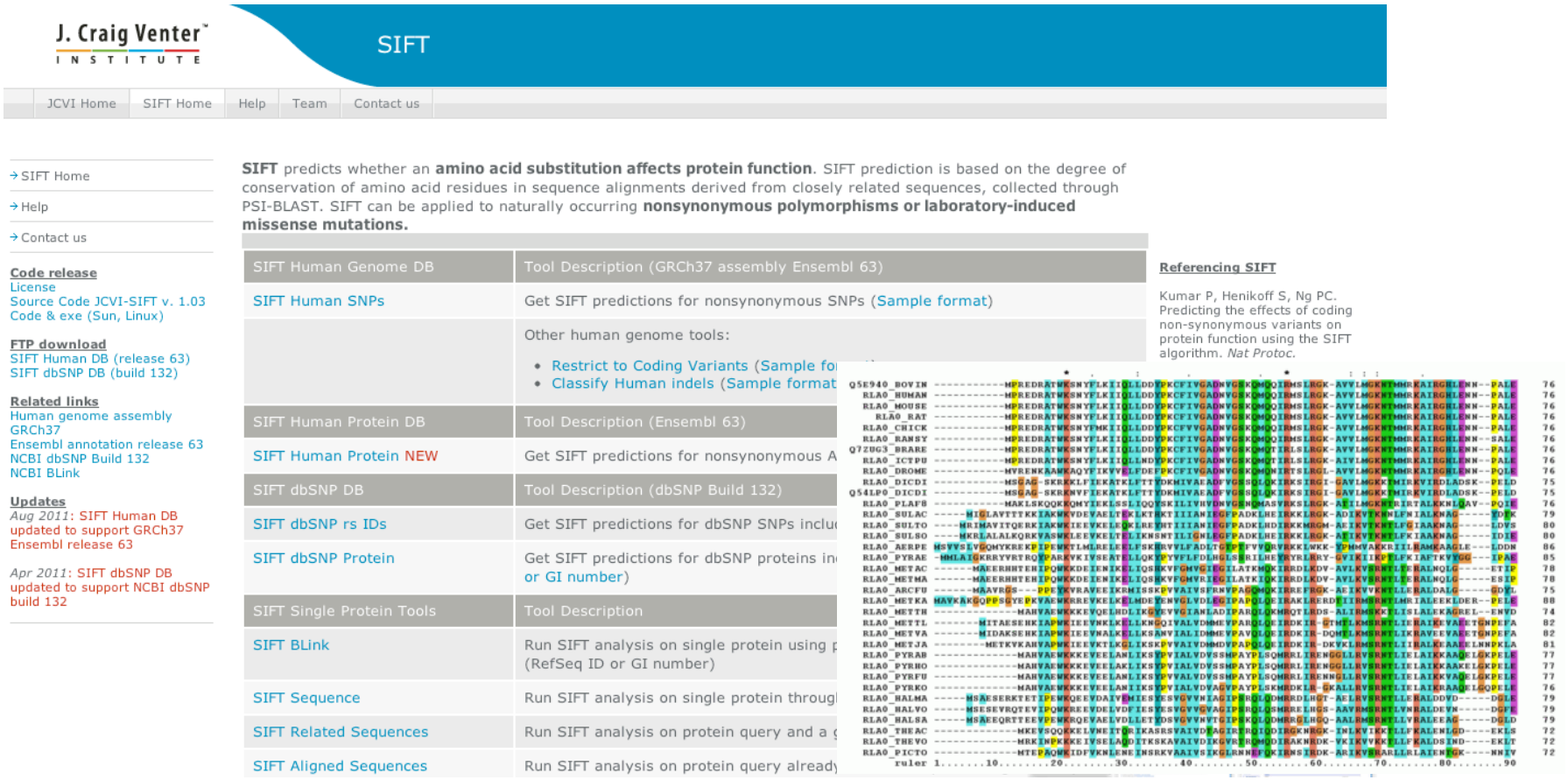


Predictors: SIFT

Predicting Deleterious Amino Acid Substitutions

Pauline C. Ng and Steven Henikoff

http://sift.jcvi.org/



- Based on the degree of conservation in a multiple sequence alignment (MSA)
- MSA generated from PSI-BLAST results (closely related sequences)
- Deleterious if SIFT ≤ 0.05

Predictors: Polyphen-2

http://genetics.bwh.harvard.edu/pph2/

A method and server for predicting damaging missense mutations

Ivan A. Adzhubei,[1,7] Steffen Schmidt,[2,7] Leonid Peshkin,[3,7] Vasily E. Ramensky,[4] Anna Gerasimova,[5] Peer Bork,[6] Alexey S. Kondrashov,[5] and Shamil R. Sunyaev[1]

MACHINE LEARNING

- Naïve Bayes Classifier

SEQUENCE BASED FEATURES

- Importance of site: DISULFID, CROSSLNK, BINDING, ACT_SITE, LIPID, METAL, SITE, MOD_RES, CARBOHYD, NON_STD…
- Importance of region: TRANSMEM, INTRAMEM, COMPBIAS, REPEAT, COILED, SIGNAL, PROPEP…
- PSIC conservation score

STRUCTURE BASED FEATURES

- Likeness to destroy hydrophobic core, electrostatic interactions, interactions with ligands, or other important features of proteins

Predictors: Polyphen-2

PolyPhen-2 prediction of functional effects of human nsSNPs

Home | About | Help | Downloads | Batch query | WHESS.db

PolyPhen-2 report for P15056 V600E

Query

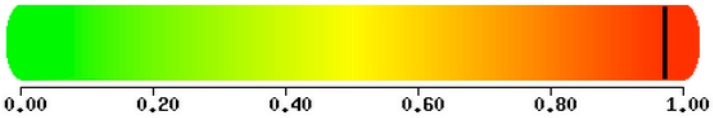
Protein Acc	Position	AA_1	AA_2	Description
P15056	600	V	E	Canonical; RecName: Full=Serine/threonine-protein kinase B-raf; EC=2.7.11.1; AltName: Full=Proto-oncogene B-Raf; AltName: Full=p94; AltName: Full=v-Raf murine sarcoma viral oncogene homolog B1; Length: 766

Results

Prediction/Confidence *PolyPhen-2 v2.2.2r398*

HumDiv

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.971** (sensitivity: **0.77**; specificity: **0.96**)



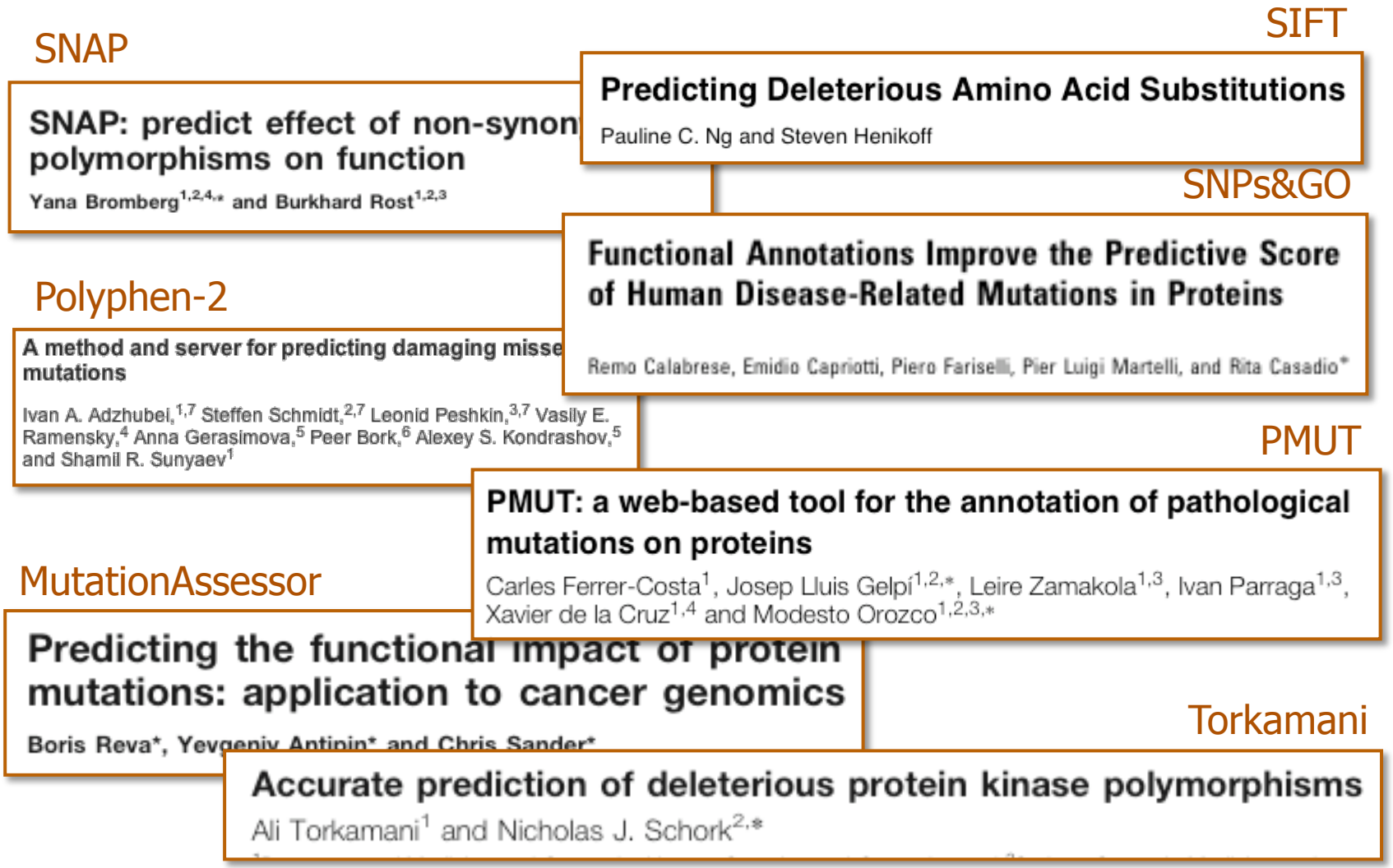
HumVar

Details

Multiple sequence alignment *UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)*

Shown are 75 amino acids surrounding the mutation position (marked with a black box). An interactive version of the complete alignment is also available.

Automatic methods to predict the pathogenicity of mutations



Some of the (many) methods implemented during the last decade

Predictors: SNPs&GO

http://snps-and-go.biocomp.unibo.it/

Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins

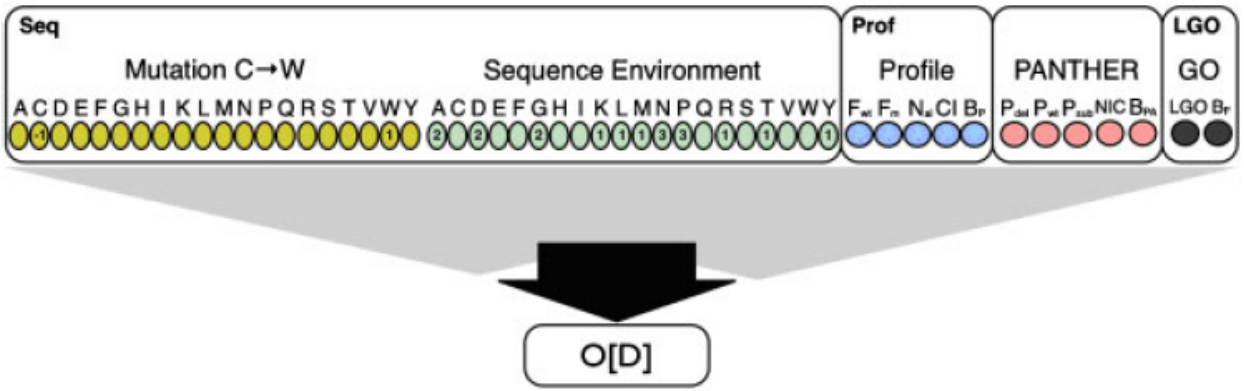
Remo Calabrese, Emidio Capriotti, Piero Fariselli, Pier Luigi Martelli, and Rita Casadio*

MACHINE LEARNING

- Support Vector Machine

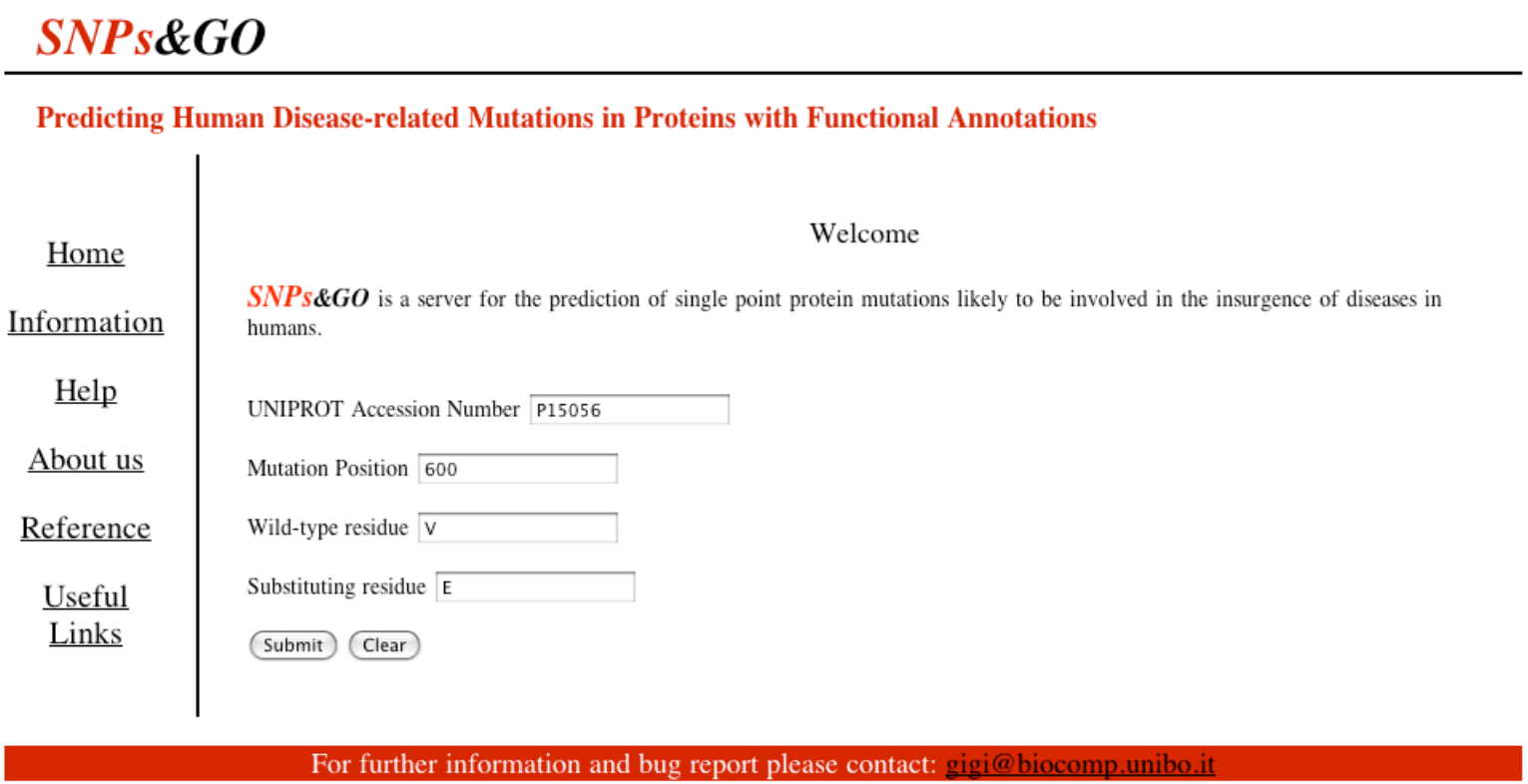
FEATURES

- Wild type and mutant amino acids
- Sequence environment (20 slots, 9-x-9)
- Profile information: Conservation in MSA (freq. of Aa, Depth, Conservation index)
- PANTHER prediction (prob. Being disease, prob. wt Aa and mut Aa in pos)
- Gene Ontology Log-odds ratio



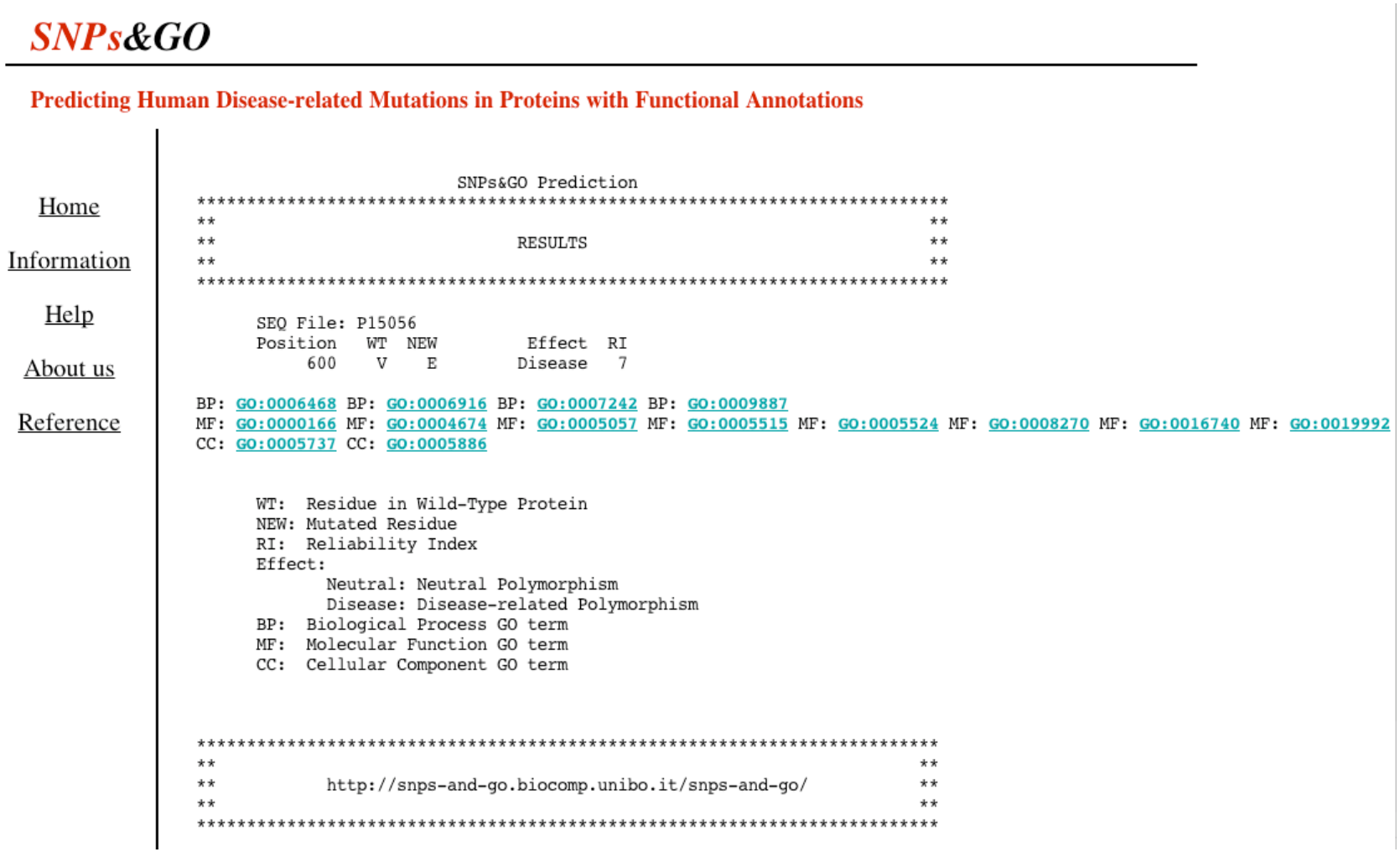
Predictors: SNPs&GO

http://snps-and-go.biocomp.unibo.it/



Predictors: SNPs&GO

http://snps-and-go.biocomp.unibo.it/



ePIPE: Prediction of protein features